DOCUMENT RESUME

ED 173 399 TM 009 441

AUTHOR Olejnik, Stephen; Porter, Andrew C.
TITLE A Longitudinal Solution to the Problem of Differential Linear Growth Patterns in Quasi-Experiments.
PUB DATE [Apr 79]
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Gains; Compensatory Education Programs; *Control Groups; Educational Research; *Experimental Groups; Predictive Measurement; Research Design; *Research Methodology; *Scoring Formulas
IDENTIFIERS *Quasi Experimental Design

ABSTRACT

Differential achievement growth patterns between comparison groups is a problem associated with data analysis in compensatory education programs. Children in greatest need of additional assistance, are usually assigned to the program rather than to an alternative treatment so that the comparison groups may vary in several ways, in addition to the effect of the treatment each group receives. A gain score procedure is recommended as a solution, using an approach that is based on the regression of achievement on time and requires the availability of pretreatment achievement data from two points in time prior to the beginning of the program. It is shown that the effect estimated is unbiased for any linear model of academic growth, including the fan spread hypothesis. Finally, the precision of the effect estimated compares favorably with competing analytic strategies. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173399

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Longitudinal Solution to the Problem of Differential Linear Growth Patterns in Quasi-Experiments

Stephen Olejnik
University of Florida

Andrew C. Porter
Michigan State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephen Olejnik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the 1979 AERA meetings in San Francisco, California

TM009 441

ABSTRACT

Differential achievement growth patterns between comparison groups has been added to the list of problems associated with data analysis in quasi-experiments. A solution to this problem is recommended in this paper. The approach suggested is based on the regression of achievement on time and requires the availability of pretreatment achievement data from two points in time prior to the beginning of the program. It is shown that the effect estimated is unbiased for any linear model of academic growth, including the fan spread hypothesis. Finally, the precision of the effect estimated compares favorably with competency analytic strategies.

The study of achievement growth patterns has become an issue of considerable interest to evaluation specialists and educational staticians. This concern with learning curves was stimulated by interest in developing new techniques for estimating the effectiveness of educational innovations. But rather than providing new solutions to old analytic problems, these efforts have only raised new questions as to the appropriateness of previously accepted estimation procedures. In particular, the growth patterns of both individuals and groups have received a great deal of attention when estimating the treatment effect of a program based on a quasi-experimental research design. Compensatory education programs are frequently evaluated using this type of research paradigm. Quasi-experiments differ from true experiments in that in the former, individuals or experimental units cannot or are not randomly assigned to alternative treatment conditions. In the case of compensatory education programs those children in greatest need of additional assistance are assigned to the program condition. As a result, the comparison groups may differ in several ways in addition to the effect of the particular treatment the groups receive. One way in which the groups may differ is in the average rate of academic achievement gained. Campbell and Boruch (1975) for example have argued that evidence supporting the theory of differential growth patterns is provided in the observation that pre-treatment achievement measures frequently indicate an initial difference in the average achievement levels of the comparison groups. This difference might be interpreted as indicating that one group has been learning at a faster rate than the other.

One example of the differential growth rate problem has been labeled the fan spread hypothesis (Campbell and Erlebacher, 1970). This theory suggests that, along with the increasing mean difference between comparison groups, there is a proportional increase in the within-group variability. Given this model of growth, traditional analysis techniques like gain scores and analysis of covariance have been challenged as inappropriate. Olejnik and Porter (1978), however, showed that in the case of analysis of covariance, the appropriateness of the technique depends on the reliability of the covariate and the nature of individual growth patterns. If the covariate is perfectly reliable or adjusted for errors of measurement and individual growth patterns are linear, then the analytic technique provides an unbiased estimate of a treatment effect. On the other hand, given a non-linear model of individual growth, analysis of covariance was shown to be inappropriate.

The fan spread model, however is only one specific example of the more general problem of differential growth rates that might be found in studies based on a quasi-experimental research design. Bryk and Weisberg (1977) have considered the more general problem by proposing several differential growth patterns, including the fan spread model. Each growth pattern studied was based on a linear model which included the average rate of group growth, a point in time when learning began and a point in time when achievement was assessed. Several analytic strategies including the analysis of covariance with a perfectly reliable covariate, gain scores and gains in standard scores were considered for each of the differential growth rate patterns. Their results indicated that except for the fan spread model, the procedures considered failed to provide for an unbiased estimate of group difference. They concluded by recommending that new data analyses strategies need to be developed which could adjust for differential growth rates.

Adjusted Gain Scores

The purpose of this paper is to suggest one possible solution to the general problem of differential linear growth patterns, including the fan spread model. The approach suggested is a gain score procedure adjusted for differences in group growth rates. Such a modification can be made if additional data collected at some time prior to the point of intervention were available. To facilitate a discussion on the development of the modified gain score procedure, figure 1 presents the achievement growth over time for a hypothetical program and control group without a treatment effect.

Insert Figure 1 about here

The horizontal axis T denotes time while the vertical axis W represents achievement. On the time dimension three points are identified: t_1, t_2 and t_3. The vertical dotted line at t_2 indicates the points of intervention while the dotted lines at t_1 and t_3 represent points in time prior to and at termination of the intervention, respectively. The solid lines represent the linear regression of achievement on time for the populations of program and control groups. In this figure the control group is shown to have a higher achievement rate (growing faster) than the program group. The points at which these growth curves intersect the dotted vertical lines represent the average achievement level on the measure administered at time t. For example (t_2, μ_{xp}) represents the population mean on measure x for the program group at the time of intervention. These solid lines can be defined in regression equations and used to predict the average group performance at any point in time. If, for example, group performance at t_3 was of interest,

the following equations may be used:

$$\mu_{yp} = a_p + b_p(t_3 - t_2)$$

$$\mu_{yc} = a_c + b_c(t_3 - t_2)$$

where:

μ_{yp}, μ_{yc} are the population mean performance on measure Y at time t, for the treatment and control groups, respectively;

a_p, a_c are the intercept constants of the growth curves at t_2 or the initial achievement levels prior to the intervention for the program and control groups respectively;

b_p, b_c are the slopes (rate of growth per unit time) of the regression line predicting achievement from time for the program and control groups respectively;

and

$t_3 - t_2$ is the period of intervention.

The difference in average performance of the program and control groups at the termination of the intervention can be determined as

(1) $$\alpha = \mu_{yp} - \mu_{yc} - (a_p - a_c) - [(b_p - b_c)(t_3 - t_2)]$$

when the intervention has no effect, the equation is as follows

$$o = \mu_{yp} - \mu_{yc} - (a_p - a_c) - [(b_p - b_c)(t_3 - t_2)]$$

The constants a_p and a_c are the initial achievement levels for the program and control groups respectively just prior to the period of intervention. The difference in previous achievement levels can be written as:

$$a_p - a_c = \mu_{xp} - \mu_{xc}$$

the difference in the mean pretest scores of the two groups. With this substitution the expression becomes the following:

$$o = (\mu_{yp} - \mu_{yc}) - (\mu_{xp} - \mu_{xc}) - [(b_p - b_c)(t_3 - t_2)]$$

The first two terms of the equation provides the estimate of a program effect if raw gain scores (post minus pre) were used. The third component (in brackets) provides an adjustment for the gain scores which is necessary under the differential growth rate condition. If the slopes are equal, that is the rate of growth is the same for both groups, the third component equals o and raw gain scores provide the appropriate adjustment procedure. This result is consistent with the conclusions drawn by Bryk and Weisberg (1977) when growth curves are parallel.

The slope of a regression line is defined as the ratio of the change in the vertical axis to the change in the horizontal axis, i.e. $b = \frac{\Delta W}{\Delta T}$. Using the information available prior to the period of intervention, the growth rate for each group can be estimated. For the program group, the regression slope can be written as the following:

$$b_p = \frac{\mu_{xp} - \mu_{zp}}{t_2 - t_1}.$$

This equation is the ratio of the change in population mean achievement at two points in time prior to the intervention with the period of time between testing. Similarly for the control group the regression slope can be written as the following:

$$b_c = \frac{\mu_{xc} - \mu_{zc}}{t_2 - t_1}$$

With these estimates of growth rates, the third component of equation 1 can be written as:

$$[(b_p - b_c)(t_3 - t_2)] = [(\frac{\mu_{xp} - \mu_{zp}}{t_2 - t_1} - \frac{\mu_{xc} - \mu_{zc}}{t_2 - t_1})(t_3 - t_2)]$$

$$= [\frac{(\mu_{xp} - \mu_{zp}) - (\mu_{xc} - \mu_{zc})}{t_2 - t_1}(t_3 - t_2)]$$

If the period of time between the first and second testing equals the period of intervention t_2 to t_3, the above equation can be simplified as the following:

$$(b_p - b_c)(t_3 - t_2) = (\mu_{xp} - \mu_{zp}) - (\mu_{xc} - \mu_{zc}) .$$

Thus, the difference in group mean gains prior to the period of intervention can be an appropriate estimate of the difference in growth rates between program and control groups. The combination of this adjustment for differential growth rates and the adjustment for differences in initial performance levels results in estimating the treatment effect as:

$$\alpha = (\mu_{yp} - \mu_{yc}) - (\mu_{xp} - \mu_{xc}) - [\frac{(\mu_{xp} - \mu_{zp}) - (\mu_{xc} - \mu_{zc})}{t_2 - t_1}(t_3 - t_2)]$$

To achieve this type of adjustment requires the creation of a new variable that takes into consideration all three test results. A variable, W, can be defined as the following:

$$W = y - x - (\bar{X} - \bar{Z}) \frac{t_3 - t_2}{t_2 - t_1}.$$

Where: x, y, $(t_3 - t_2)$, $(t_2 - t_1)$ are as defined earlier;

$\bar{X}$ is the group mean on the pretest administered at t_2 or the point of intervention; and

$\bar{Z}$ is the group mean on the pretest administered at t_1, or some point in time prior to the intervention.

If the time between the first two test administrations equals the period of intervention, the new variable is simply a gain minus the difference of the group's means at two points in time prior to the intervention. This second factor adjusts for the growth rate of the group.

This solution to the problem of differential growth rates can be related to Bryk and Weisberg's (1977) analysis of the issue. In their analyses individual achievement was defined as:

$$\text{Achievement} = \pi_i(t_j - \tau_g).$$

where:

π_i is the growth rate for individual (i);

τ_i is the point in time when individual (i) begins to learn; and

t_j is the point in time (j) when achievement is assessed.

The average achievement for a particular group was then defined as:

$$\text{Average Achievement} = \mu_\pi(t_j - \mu_\tau).$$

where:

μ_π is the average growth rate of the group;

μ_τ is the average point in time when individuals within the group began to learn; and

t_j is as defined above.

Based on this simple model of linear growth, the authors show that the appropriate adjustment for all linear patterns of growth is:

$$\frac{\mu_{\pi p}(t_3 - \mu_{\tau p}) - \mu_{\pi c}(t_3 - \mu_{\tau c})}{\mu_{\pi c}(t_2 - \mu_{\tau p}) - \mu_{\pi c}(t_2 - \mu_{\tau c})} \quad (\mu_{xp} - \mu_{xc})$$

The adjusted gain score procedure suggested in the present paper can be shown to be identical to Bryk and Weisberg's solution when group growth rates are estimated using the data from the two pretests:

$$b_p = \mu_{\pi p} = \frac{\mu_{xp} - \mu_{zp}}{t_2 - t_1} \quad \text{and} \quad b_c = \mu_{\pi c} = \frac{\mu_{xc} - \mu_{zc}}{t_2 - t_1} .$$

That is:

$$\frac{\mu_{\pi p}(t_3 - \mu_{\tau p}) - \mu_{\pi c}(t_3 - \mu_{\tau c})}{\mu_{\pi p}(t_2 - \mu_{\tau}) - \mu_{\pi c}(t_2 - \mu_{\tau c})} (\mu_{xp} - \mu_{xc}) = (\mu_{xp} - \mu_{xc}) + [(\mu_{xp} - \mu_{zp}) - (\mu_{xc} - \mu_{zc})]$$

when the period of time between the two pretests equals the period of intervention. (See Appendix A)

Thus the adjusted gain score strategy provides an appropriate solution to all differential linear growth rate patterns including the fan spread hypothesis. It might be noted that individual growth patterns need not be linear for the average rate of growth for the group to be linear.

The Standard Error of the Adjusted Gain Scores

The above discussion has shown that the adjusted gain score procedure provides an unbiased estimate of a treatment effect in situations where comparison groups are growing at differential rates. To test for statistical significance and to judge the precision of the estimate the standard error of the difference between the comparison group means on the adjusted variable is needed. The adjusted gain score strategy suggested that an individual's post treatment achievement score be adjusted for the initial pretreatment achievement level and an estimate of group's growth rate:

$$W = y - x - (\bar{x} - \bar{z}) \quad \text{or}$$

$$W = y - x - \bar{x} + \bar{z}$$

The variance of this variable can be written as the following:

$$\text{Var}(W) = \text{Var}(y - x - \bar{x} + \bar{z}) = \text{Var}(y) + \text{Var}(x) + \text{Var}(\bar{x}) + \text{Var}(\bar{z})$$
$$- 2\,\text{Cov}(y,x) - 2\,\text{Cov}(y,\bar{x}) + 2\,\text{Cov}(y,\bar{z}) + 2\,\text{Cov}(x,\bar{x})$$
$$- 2\,\text{Cov}(x,\bar{z}) - 2\,\text{Cov}(\bar{x},\bar{z})$$

Which can be simplified to: (See Appendix B)

$$\text{Var}(W) = \text{Var}(y) + (1 + \tfrac{3}{n})\,\text{Var}(x) - (2 + \tfrac{2}{n})\,\text{Cov}(x,y)$$
$$+ \tfrac{1}{n}[\text{Var}(z) + 2\,\text{Cov}(y,z) - 4\,\text{Cov}(x,z)]$$

The contrast of interest is in terms of the group means on the adjusted variable $\bar{W}$ resulting in the following statement for the standard error:

$$\sqrt{\tfrac{2}{n}[\text{Var}(y) + (1 + \tfrac{3}{n})\,\text{Var}(x) - (2 + \tfrac{2}{n})\,\text{Cov}(x,y)] + \tfrac{2}{n^2}[\text{Var}(z) + 2\text{Cov}(y,z) - 4\text{Cov}(x,z)]}$$

As the sample size increases the contribution the terms in the second bracket make to the magnitude of the standard error becomes minimal. Olejnik (1977) compared the above standard error with the standard errors for gains in standard scores and analysis of covariance adjusted for errors of measurements in the covariate. His results indicated that in small samples the standard error for adjusted gain scores was approximately same as that for the two competing strategies. With large samples however the standard error of the adjusted gain score was considerably less than either gains in standard scores or analysis of covariance with estimated true scores.

One group vs two group research designs

The adjusted gain score strategy has been suggested as an appropriate analytic technique where groups are changing linearly. Given the assumptions of linear growth and the availability of data from two pretests, the necessity of having an independent control group may be questioned. Under these conditions it might be argued that a treatment group could serve as its own control. This could be achieved by defining a regression line based on the pre-intervention data and predicting posttest scores under a no-treatment effect condition. Differences between the obtained posttest performances and the predicted performances might then be attributed to the treatment. While this approach might be used, it would be difficult to argue that differences between the observed and expected posttest performances were due solely to the treatment. History, maturation, testing and regression effects are all reasonable threats to internal validity for this approach. The use of an independent control group provides a means by which the effects of these threats to internal validity can be controlled as long as those effects are equal across the two groups.

In addition to the issue of internal validity, a further consideration is the question of precision. With an independent control group, the contrast of interest is the difference between the adjusted group means. The standard error associated with that contrast can be written as

$$\sqrt{2\,\mathrm{Var}(\bar{W})}\quad .$$

The degrees of freedom associated with this test is the sum of the two sample sizes minus 2. In the single group design the contrast of interest is the adjusted group mean minus zero. Its standard error can be written as

$$\sqrt{\mathrm{Var}\ (\bar{W})}$$

The degrees of freedom associated with this test is the number of individuals in the group minus one. Comparing the one group versus the two group design in terms of precision would indicate smaller error variance associated with the former approach. The two group design, however, has more degrees of freedom. The difference between the degrees of freedom becomes negligible, however, when the sample is large. Thus, the one group design is likely to provide greater precision than the two group design. The problems related to threats to internal validity in the one group design however, discourage the use of this approach.

References

Bryk, A. S. & Weisberg, H. Use of the non-equivalent control group design when subjects are growing. Psychological Bulletin, 1977.

Campbell, D. T. & Boruch, R. F. Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. A. Bennett (Ed.), Evaluation and experiment. New York: Academic Press, 1975.

Campbell, D. T. & Erlebacher, A. How regression artifacts in quasi-experimantal evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Compensatory education: A national debate (Vol. 3). The Disadvantaged Child. New York: Brunner/Mazel, 1970.

Olejnik, S. Data analysis strategies for quasi-experimental studies where differential group and individual growth rates are assumed. Unpublished Ph.D. dissertation, East Lansing: Michigan State University, 1974.

Olejnik, S. & Porter, A. C. Estimating treatment effects and precision for quasi-experiments assuming differential group and individual growth patterns. Paper presented at the 1978 American Educational Research Association meeting in Toronto, Canada.

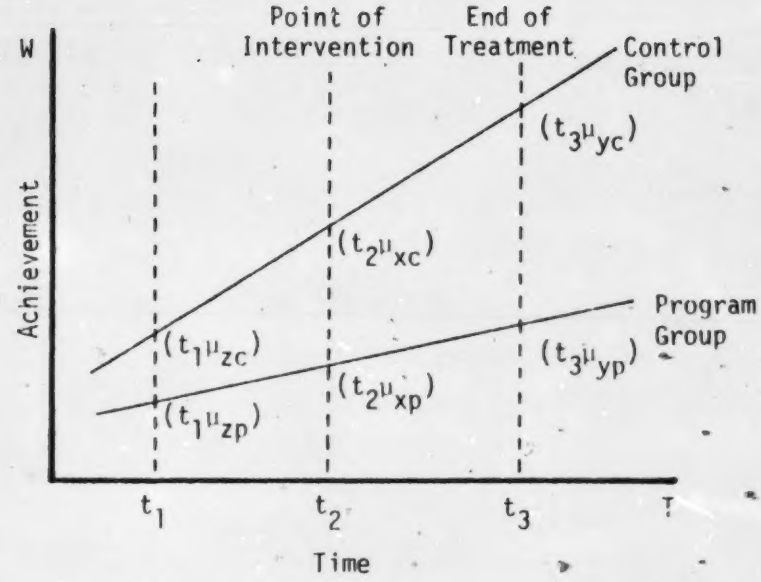

Figure 1. Differential growth rates considered over three points in time.

APPENDIX A

The equivalence of adjusted gain scores with the adjustment factor suggested by Bryk and Weisberg (1977).

$$(\mu_{xp} - \mu_{xc}) + [(\mu_{xp} - \mu_{zp}) - (\mu_{xc} - \mu_{zc})] = (\mu_{xp} - \mu_{xc}) \frac{\mu_{\pi p}(t_3 - \mu_{\tau p}) - \mu_{\pi c}(t_3 - \mu_{\tau c})}{\mu_{\pi p}(t_2 - \mu_{\tau p}) - \mu_{\pi c}(t_2 - \mu_{\tau c})}$$

Let

$\mu_{\pi p}$, $\mu_{\pi c}$ represent the growth rates for the program and control groups respectively;

$\mu_{\tau p}$, $\mu_{\tau c}$ represent the initial learning point for the program and control groups respectively;

t_1, t_2, t_3 represent points in time when achievement is assessed.

$\mu_{zp}, \mu_{xp}, \mu_{yp}$ represent the average achievement of the program group on the z, x and y achievement measures.

$\mu_{zc}, \mu_{xc}, \mu_{yc}$ represent the average achievement of the program group on the z, x and y achievement measures.

Bryk and Weisberg suggest the following adjustment factor for any linear model of group growth:

$$\frac{\mu_{\pi p}(t_3 - \mu_{\tau p}) - \mu_{\pi c}(t_3 - \mu_{\tau c})}{\mu_{\pi p}(t_2 - \mu_{\tau p}) - \mu_{\pi c}(t_2 - \mu_{\tau c})} (\mu_{xp} - \mu_{xc})$$

Adjusted gains scores suggest that rate of growth can be estimated as:

(1) $$\mu_{\pi p} = \frac{\mu_{xp} - \mu_{zp}}{t_2 - t_1}, \quad \mu_{\pi c} = \frac{\mu_{xc} - \mu_{zc}}{t_2 - t_1}.$$

Substituting these estimates in Bryk and Weisbergs's adjustment factor results in the following:

$$\frac{\frac{1}{t_2 - t_1}[(\mu_{xp} - \mu_{zp})(t_3 - \mu_{\tau p}) - (\mu_{xc} - \mu_{zc})(t_3 - \mu_{\tau c})]}{\frac{1}{t_2\ t_1}[(\mu_{xp} - \mu_{zp})(t_2 - \mu_{\tau p}) - (\mu_{xc} - \mu_{zc})(t_2 - \mu_{\tau c})]} (\mu_{xp} - \mu_{xc})$$

$$\frac{(\mu_{xp} - \mu_{zp})[(t_2 - \mu_{\tau p}) + (t_3 - t_2)] - (\mu_{xc} - \mu_{zc})[(t_2 - \mu_{\tau c}) + (t_3 - t_2)]}{(\mu_{xp} - \mu_{zp})(t_2 - \mu_{\tau p}) - (\mu_{xp} - \mu_{zp})(t_2 - \mu_{\tau c})} (\mu_{xp} - \mu_{xc})$$

$$\left[\frac{(\mu_{xp} - \mu_{zp})(t_3 - t_2) - (\mu_{xc} - \mu_{zc})(t_3 - t_2)}{(\mu_{xp} - \mu_{zp})(t_2 - \mu_{\tau p}) - (\mu_{xp} - \mu_{zp})(t_2 - \mu_{\tau c})} + 1 \right] (\mu_{xp} - \mu_{xc})$$

$$\frac{(\mu_{xp} - \mu_{xc})(t_3 - t_2)[(\mu_{xp} - \mu_{zp}) - (\mu_{xc} - \mu_{zc})]}{(\mu_{xp} - \mu_{zp})(t_x - \mu_{\tau p}) - (\mu_{xc} - \mu_{zc})(t_2 - \mu_{\tau c})} + (\mu_{xp} - \mu_{xc})$$

From (1) above:

(2) $\mu_{xp} - \mu_{zp} = \mu_{\pi p}(t_2 - t_1)$ and

$\mu_{xc} - \mu_{zc} = \mu_{\pi c}(t_2 - t_1)$

If the period of time between the pretests equals the period of intervention $(t_2 - t_1) = (t_3 - t_2)$ then

(3) $$t_3 - t_2 = \frac{(\mu_{xp} - \mu_{zp})}{\mu_{\pi p}}$$

Furthermore from Bryk and Weisberg's definition of group growth

(4) $\mu_{xp} = \mu_{\pi p}(t_2 - \mu_{\tau p})$

$$t_2 - \mu_{\tau p} = \frac{\mu_{xp}}{\mu_{\pi p}}$$

Then combining the results of 2, 3 and 4

$\mu_{xp}(t_3 - t_2) = (\mu_{xp} - \mu_{zp})(t_2 - \mu_{\tau p})$ and

$\mu_{xc}(t_3 - t_2) = (\mu_{xc} - \mu_{zc})(t_2 - \mu_{\tau c})$

Thus $(\mu_{xp} - \mu_{xc})(t_3 - t_2) = (\mu_{xp} - \mu_{zp})(t_2 - \mu_{\tau p}) - (\mu_{xc} - \mu_{zc})(t_2 - \mu_{\tau c})$

and:

$$[(\mu_{xp} - \mu_{zp}) - (\mu_{xc} - \mu_{zc})] + (\mu_{xp} - \mu_{xc}) = \frac{\mu_{\pi p}(t_3 - \mu_{\tau p}) - \mu_{\pi c}(t_3 - \mu_{\tau c})}{\mu_{\pi p}(t_2 - \mu_{\tau p}) - \mu_{\pi c}(t_2 - \mu_{\tau c})} (\mu_{xp} - \mu_{xc})$$

APPENDIX B

THE VARIANCE OF THE ADJUSTED GAIN SCORE VARIABLE

Let: Z represent the first pretest

X represent the pretest administered immediately prior to the intervention

Y represent the post-intervention test performance

Then the adjusted score is determined as follows:

$$W = Y - X - [\bar{X} - \bar{Z}]$$

$$W = Y - X - \bar{X} + \bar{Z}$$

$$\text{Var}(W) = \text{Var}[Y - X - \bar{X} + \bar{Z}]$$

$$= \text{Var}(Y) + \text{Var}(X) + \text{Var}(\bar{X}) + \text{Var}(\bar{Z})$$

$$- 2\text{Cov}(Y,X) - 2\text{Cov}(Y,\bar{X}) + 2\text{Cov}(Y,\bar{Z}) + 2\text{Cov}(X,\bar{X}) - 2\text{Cov}(X,\bar{Z}) - 2\text{Cov}(\bar{X},\bar{Z})$$

$$\text{Cov}(Y,\bar{X}) = \frac{\sum_{N}\sum_{n}(Y - \mu_y)(\bar{X} - \mu_x)}{Nn}$$

$$= \sum_{N}\sum_{n} Y\bar{X} - \sum_{N}\sum_{n} Y\mu_x - \sum_{N}\sum_{n} \bar{X}\mu_y + \sum_{N}\sum_{n} \mu_y\mu_x/Nn$$

$$= \sum_{N} n\bar{Y}\bar{X} - \sum_{N} n\bar{Y}\mu_x - Nn\,\mu_x\mu_y + Nn\,\mu_y\mu_x/Nn$$

$$= \sum_{N} \bar{Y}\bar{X} - \mu_x \sum_{N} \bar{Y}/N$$

$$= \frac{\sum_{N} \bar{Y}\bar{X} - \frac{\mu_x\,\mu_y}{N}}{N}$$

$$= \text{Cov}(\bar{Y},\bar{X})$$

$$\text{Cov}\ (Y,\bar{Z}) = \text{Cov}\ (\bar{Y},\bar{Z})$$

$$\text{Cov}\ (X,\bar{X}) = \text{Var}\ (\bar{X})$$

$$\text{Cov}\ (X,\bar{Z}) = \text{Cov}\ (\bar{X},\bar{Z})$$

Assume a single population with observations on two variables X and Y. Select N random samples of size n from that population, then:

$$\text{Cov}\ (\bar{X},\bar{Y}) = \frac{\text{Cov}\ (X,Y)}{n}$$

$$\text{Cov}\ (\bar{X},\bar{Y}) = \frac{\sum^{N} (\bar{X} - \mu_x)(\bar{Y} - \mu_y)}{N}$$

$$= \sum^{N}(\bar{X}\,\bar{Y} - \bar{X}\,\mu_y - \bar{Y}\,\mu_x + \mu_x\,\mu_y)/N$$

$$= \left(\sum^{N} \bar{X}\,\bar{Y} - \sum^{N} \bar{X}\,\mu_y - \sum^{N} \bar{Y}\,\mu_x + \sum^{N} \mu_x\,\mu_y\right)/N$$

$$= (\sum^{N} \bar{X}\,\bar{Y} - N\,\mu_x\,\mu_y - N\,\mu_y\,\mu_x + N\,\mu_x\,\mu_y)/N$$

$$= \left(\sum^{N} \bar{X}\,\bar{Y} - N\,\mu_x\,\mu_y\right)/N$$

$$\sum^{N} \bar{X}\,\bar{Y} = \sum^{N} \frac{\sum^{n} X \sum^{n} Y}{n^2}$$

$$\sum^{n} X \sum^{n} Y = \sum^{n} X_i Y_i + n(n-1)\,\mu_x\,\mu_y$$

i.e., $(X_1 + X_2)(Y_1 + Y_2) = X_1Y_1 + X_2Y_2 + X_1Y_2 + X_2Y$

assuming that all pairs where $i \neq i'$ are independent and there are $n(n-1)$ such pairs.

$$\frac{1}{n}\sum^{N}\left(\sum^{n} X \sum^{n} Y\right) = \frac{1}{n^2}\left[\sum^{N}\sum^{n}\sum^{n} X Y - \sum^{N} n\,\mu_x\,\mu_y\right] + \frac{1}{n^2}\sum^{N} n^2\,\mu_x\,\mu_y$$

$$\left[\frac{1}{n^2}\left[\sum^{N}\sum^{n}\sum^{n} X Y - Nn\,\mu_x\,\mu_y\right] + N\,\mu_x\,\mu_y - N\,\mu_x\,\mu_y\right] / N$$

$$\frac{1}{n}\left[\sum^{N}\sum^{n}\sum^{n} X Y - N n\,\mu_x\,\mu_y\right] / Nn$$

$$= \frac{\frac{1}{n}\left[\sum^{N}\sum^{n}\sum^{n} X Y - \frac{\sum^{N}\sum^{n} X \sum^{N}\sum^{n} Y}{Nn}\right]}{Nn}$$

$$= \frac{1}{n}\,\mathrm{Cov}\,(X\,Y).$$

$$\mathrm{Var}\,(W) = \mathrm{Var}(Y) + \mathrm{Var}(X) + \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Z}) - 2\mathrm{Cov}(Y,X) - 2\mathrm{Cov}(\bar{Y},\bar{X}) + 2\mathrm{Cov}(\bar{Y},\bar{Z})$$

$$+ 2\mathrm{Var}(\bar{X}) - 2\mathrm{Cov}(\bar{X},\bar{Z}) - 2\mathrm{Cov}\,(\bar{X},\bar{Z})$$

Assuming independent samples $\mathrm{Cov}(\bar{X},\bar{Y}) = \frac{\mathrm{Cov}(X,Y)}{n}$.

$$\text{Var}(W) = \text{Var}(Y) + \text{Var}(X) + \frac{1}{n}\text{Var}(X) + \frac{1}{n}\text{Var}(Z) - 2\text{Cov}(Y,X) - \frac{2}{n}\text{Cov}(Y,X)$$

$$+ \frac{2}{n}\text{Cov}(Y,Z) + \frac{2}{n}\text{Var}(X) - \frac{4}{n}\text{Cov}(X,Z)$$

$$= \text{Var}(Y) + \left[\text{Var}(X) + \frac{1}{n}\text{Var}(X) + \frac{2}{n}\text{Var}(X)\right] + \left[-2\,\text{Cov}(X,Y) - \frac{2}{n}\text{Cov}(X,Y)\right]$$

$$+ \frac{1}{n}\text{Var}(Z) + \frac{2}{n}\text{Cov}\,(Y,Z) - \frac{4}{n}\text{Cov}(X,Z)$$

$$= \text{Var}(Y) + \left(1 + \frac{3}{n}\right)\text{Var}(X) - \left(2 + \frac{2}{n}\right)\text{Cov}(X,Y)$$

$$+ \frac{1}{n}\left[\text{Var}(Z) + 2\text{Cov}(Y,Z) - 4\text{Cov}(X,Z) - 4\text{Cov}(x,Z)\right]$$

$$\text{Var}(\bar{W}) = \frac{1}{n}\left[\text{Var}(Y) + \left(1 + \frac{3}{n}\right)\text{Var}(X) - \left(2 + \frac{2}{n}\right)\text{Cov}(X,Y)\right]$$

$$+ \frac{1}{n^2}\left[\text{Var}(Z) + 2\text{Cov}(Y,Z) - 4\text{Cov}(X,Z)\right].$$